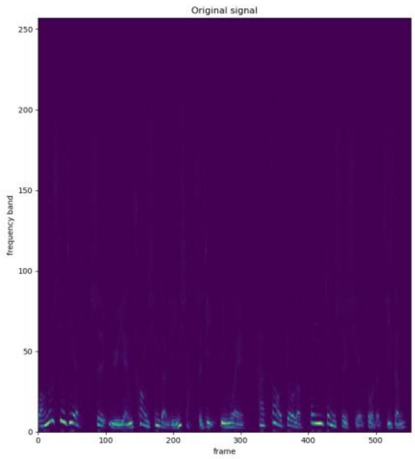
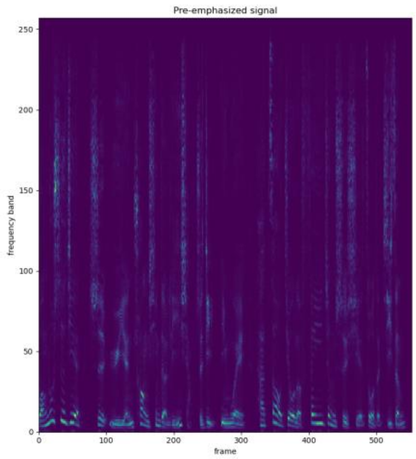
# DSP Lab01 Audio & Speech: MFCC

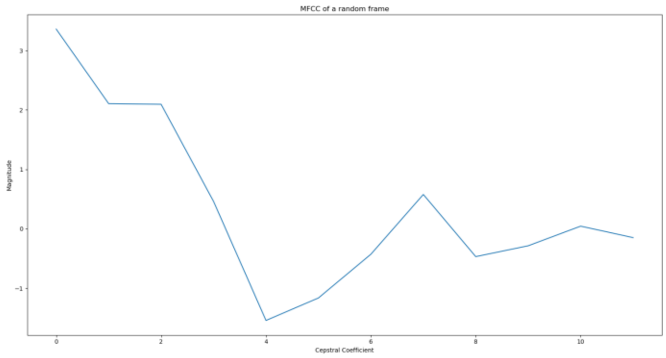
## 108061129 陳楷芮

1. Demo (1)

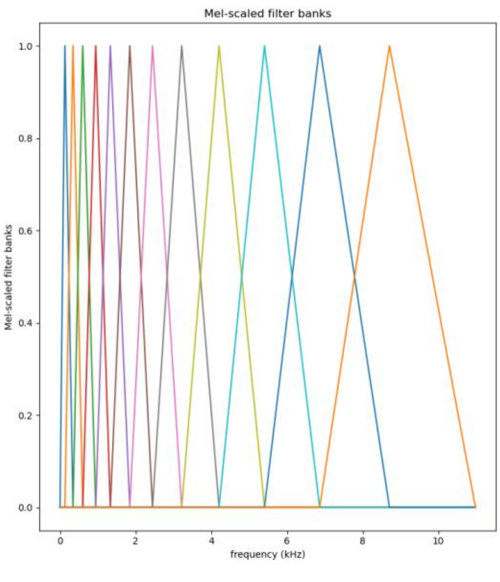


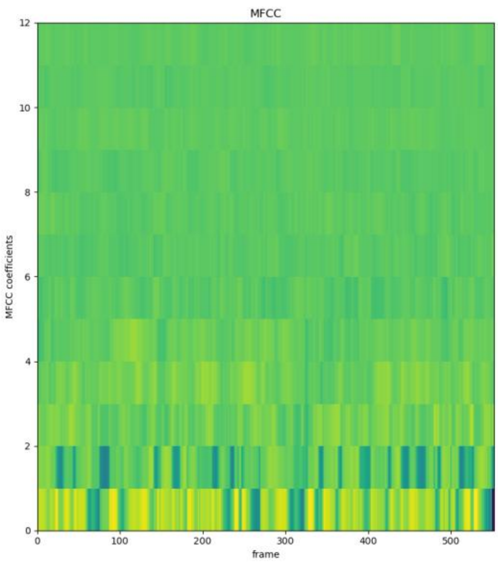




2. Demo (2)




3. Demo (3)



## Report Questions

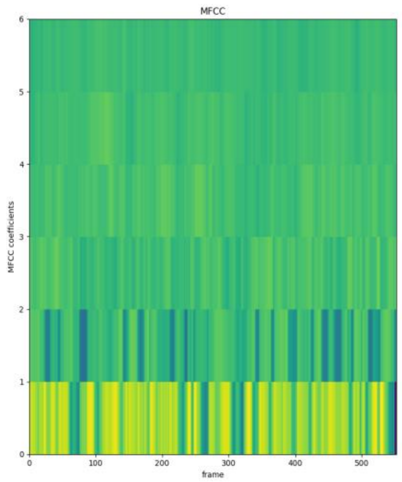
1. **Give an intuitive and a mathematical explanation of time-domain pre-emphasis; suggest why this step is beneficial for audio analysis?**

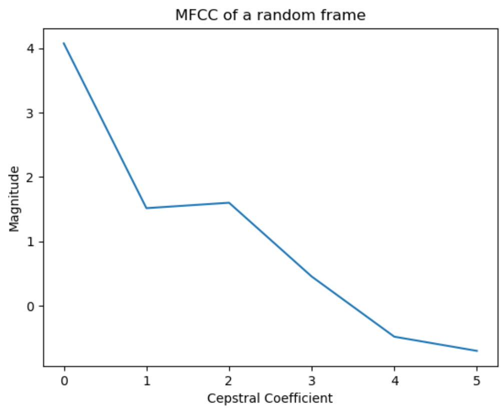
   As we can see from the formula of pre-emphasis, $H(z)=1-a*z-1$, since there is a zero, meaning that it's an HPF. The reason why we need an HPF is below.

   Intuitively, since the mechanism of human sound production suggests that there are some low-frequency noises produced during the process, therefore, the high-frequency part is suppressed. To compensate the suppressed high-frequency sound and suppress the unwanted low-frequency noise, HPF is needed.

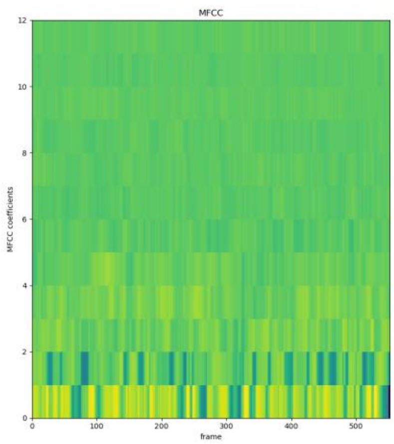
2. **Compare the MFCC of a chosen frame under different number of banks/coefficients. Compare the MFCC heatmap under different number of banks/coefficients. Is it the more the merrier? Why or why not? In case of this audio file, what is the number of banks you would choose? Why?**

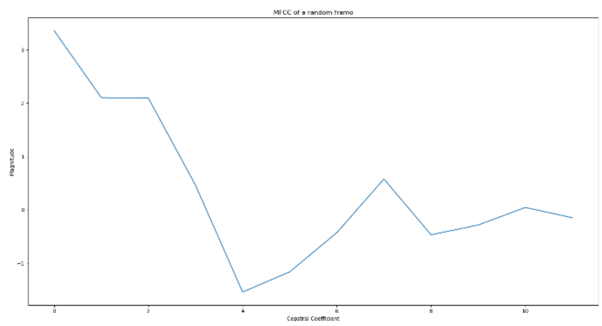
   (1) Number of banks: 6



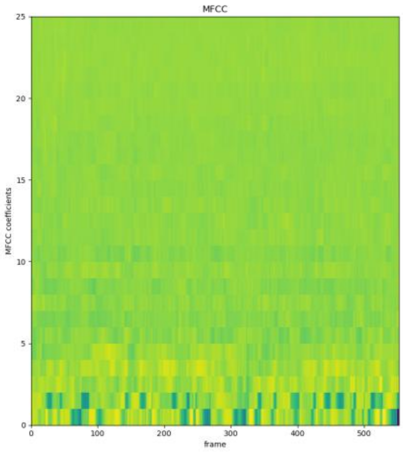



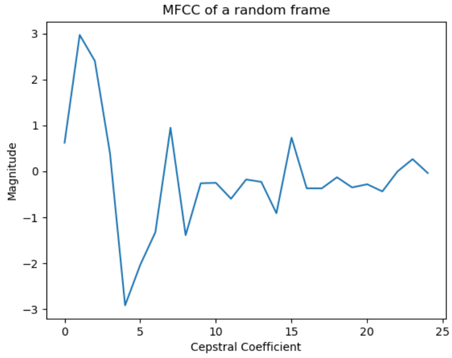

   (2) Number of banks: 12



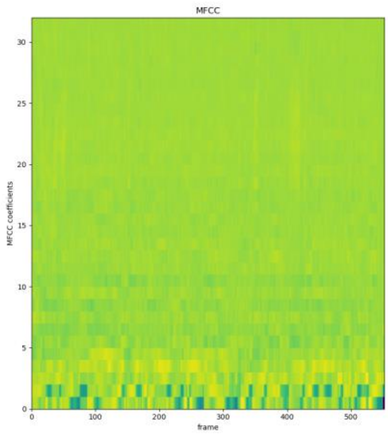



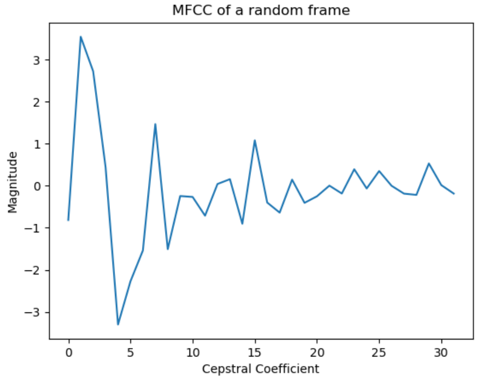
(3) Number of banks: 25




(4) Number of banks: 32

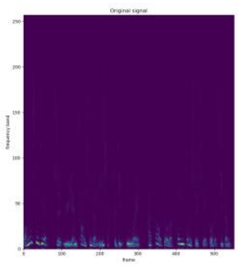


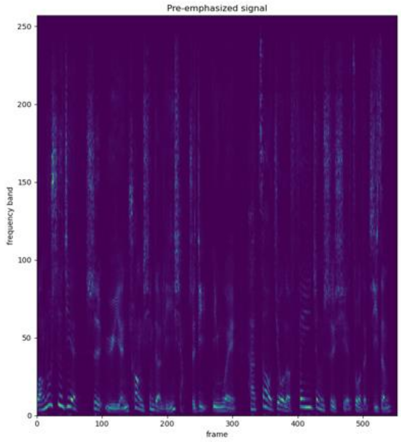

As we can see from above, the optimal number of banks is roughly about 25， since from 12 to 25, the changes in the heatmap is still significant, however, from 25 to 32, it doesn't change that much compare to the former range. Furthermore, when increase a bank, the data size also increases, causing calculating problems. Therefore, the more of the banks isn't the merrier, there's a tradeoff between quality and data size.

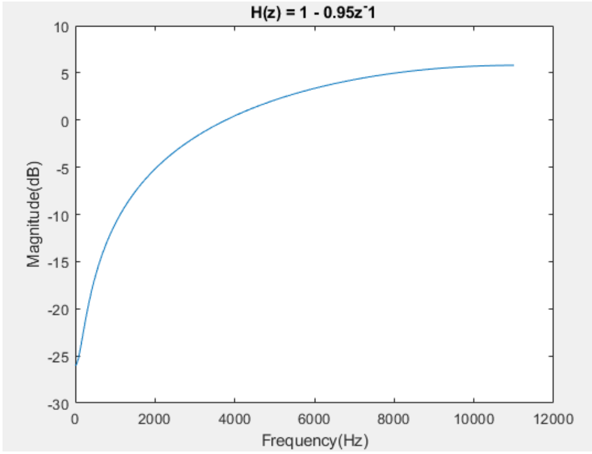
## Bonus Questions

1. **Plot the magnitude curve of H(z) with coefficients 0.95, 0.99, 0.65 in MATLAB. Change the pre-emphasis coefficient in your lab code correspondingly and observe the difference on the spectrogram. How does the coefficient influence the results?**

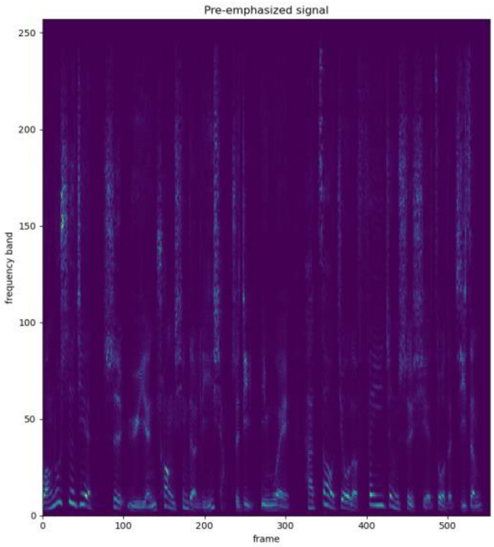


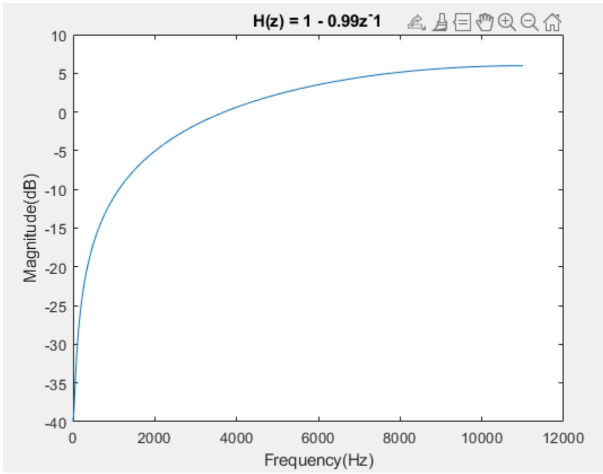
(1) When coefficient: 0.95



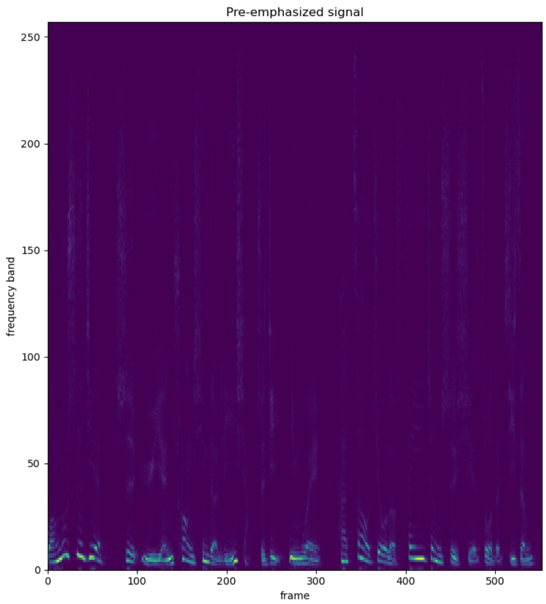



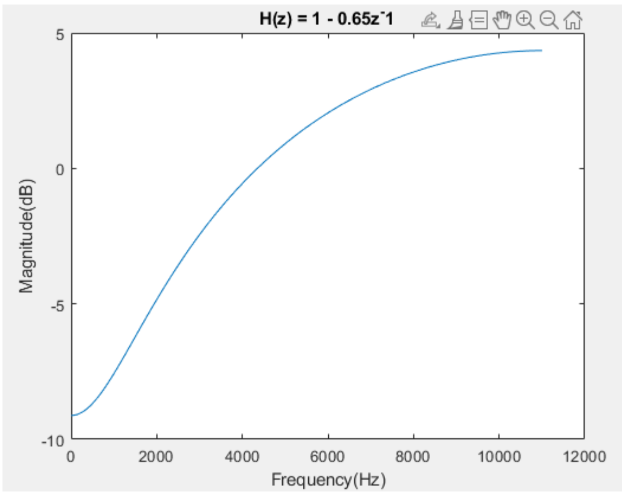

(2) When coefficient: 0.99




(3) When coefficient: 0.65



From the above, we can see that as the coefficient increases, it attenuates the low-frequency part and enhance the high frequency part.

2. **Notice in the code there is a discrepancy between our chosen number of FFT frequency quantization and the number of frequency channels in the output figure. Explain why the number of frequency channels halved and plot the spectrogram in which you perform conventional FFT.**

To analyze the sound efficiently, we partitioned the sound into frames, in this case, a frame includes 512 samples. However, to avoid two adjacent frames be too far from each other inducing data loss, there must be some overlapping between two adjacent frames. If we only consider the overlapped part, there are about half number of the frames, causing the result of the halved channels.

If we perform conventional FFT, the corresponding spectrogram is below.